Mining the Genome for Therapeutic Targets

Jose C. Florez

Current pharmacological options for type 2 diabetes do not cure the disease. Despite the availability of multiple drug classes that modulate glycemia effectively and minimize long-term complications, these agents do not reverse pathogenesis, and in practice they are not selected to correct the molecular profile specific to the patient. Pharmaceutical companies find drug development programs increasingly costly and burdensome, and many promising compounds fail before launch to market. Human genetics can help advance the therapeutic enterprise. Genomic discovery that is agnostic to preexisting knowledge has uncovered dozens of loci that influence glycemic dysregulation. Physiological investigation has begun to define disease subtypes, clarifying heterogeneity and suggesting molecular pathways for intervention. Convincing genetic associations have paved the way for the identification of effector transcripts that underlie the phenotype, and genetic or experimental proof of gain or loss of function in select cases has clarified the direction of effect to guide therapeutic development. Genetic studies can also examine off-target effects and furnish causal inference. As this information is curated and made widely available to all stakeholders, it is hoped that it will enhance therapeutic development pipelines by accelerating efficiency, maximizing cost-effectiveness, and raising ultimate success rates.

STATEMENT OF THE PROBLEM

The current state of affairs is deeply unsatisfying. Despite its status as one of the oldest documented endocrinopathies, the availability of a molecular therapy for almost a century, and the substantial morbidity and mortality that make type 2 diabetes a modern urgent public health menace, we have not been able to cure the disease—at least, using pharmacological means. Our surgical colleagues have had to lead the way by demonstrating restoration of euglycemia after gastric bypass leading to lasting remissions, an outcome that is also achievable (but much more challenging) through behavioral means, if adopted and adhered to early in the disease course. But the academic community and the pharmaceutical industry have been unable to produce a drug that reverses pathophysiology and permanently rescues an individual from type 2 diabetes.

Although it is true that drug discovery has led to a proliferation of drug classes targeting this condition (1) (with 12 drug classes approved by the U.S. Food and Drug Administration at last count, including insulin and its analogs, biguanides, sulfonylureas, α-glucosidase inhibitors, thiazolidinediones, glinides, glucagon-like peptide 1 [GLP-1] receptor agonists, dipeptidyl peptidase 4 inhibitors, bile acid resins, dopamine agonists, pramlintide, and sodium–glucose cotransporter 2 [SGLT2] inhibitors), the majority of these agents simply address a by-product of the disease process: they are symptom treating, but not disease modifying. Merely lowering glucose by interfering with its gastrointestinal absorption, reducing its hepatic release, enhancing its uptake into insulin-responsive tissues, or favoring its renal elimination does little to modify the pathogenic insults that cause primary β-cell degeneration or target-organ insulin resistance (2). Modulating glycemia is indeed crucial to reduce long-standing microvascular and even macrovascular complications, but we are stuck in secondary prevention rather than in cure mode.

In part, this is because of our limited understanding of disease pathogenesis. Diabetes is defined by a diagnostic metric (hyperglycemia) that only reflects the end result of many altered processes. A patient who develops hyperglycemia in the absence of autoimmunity and without a clear inherited pattern of transmission is automatically given the diagnosis of type 2 diabetes and entered into a treatment algorithm that does not address the molecular causes of his or her current metabolic state, let alone

Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital; Programs in Metabolism and Medical and Population Genetics, Broad Institute; and Department of Medicine, Harvard Medical School, Boston, MA

Corresponding author: Jose C. Florez, jcflorez@mgh.harvard.edu.

make an attempt to tailor therapies to specific pathways (3). This is akin to defining cancer on the sole basis of mass effect on surrounding anatomical structures and instituting nonspecific therapies to control tissue growth, a paradigm that thankfully has been superseded throughout most of oncological practice. There is little doubt that type 2 diabetes is a conglomerate of multiple pathophysiological derangements with variable manifestations in a given patient; thus, there is a peremptory need to elucidate its heterogeneity and explore whether we can classify the disease in physiologically driven, clinically relevant subtypes (4).

Two other obstacles stand in the way of novel pharmacological therapeutics for type 2 diabetes (Table 1). First is the inordinate cost of drug development (5). In the current era, it is not unusual for a drug program to incur expenses of over $1 billion to go from molecule discovery to market, severely curtailing new compound development to only the best capitalized companies, inhibiting risk-taking around new chemical entities, and undermining innovation (6). In type 2 diabetes, costs have been magnified since the U.S. Food and Drug Administration began requiring proof of cardiovascular safety for new type 2 diabetes agents, necessitating the conduct of long and expensive cardiovascular clinical trials. As a consequence, in 2013 57.6% of all diabetes expenditures in the U.S. ($101.4 billion) went to pharmaceuticals (7). Second, and connected with the above, we have to contend with the dismal and declining success rate of many drug programs, with only about 10% of drugs ever making it to market through failure to show efficacy or preserve safety in humans (8). Thus the cost of successful medications in part subsidizes the many failed attempts elsewhere in the drug pipeline (9).

Table 1—Challenges to drug development in type 2 diabetes

Unclear heterogeneity of the disease
    Type 2 diabetes is used as a "catch-all" diagnosis.
    Metabolic state changes with disease progression.
    Disease subclassification is not routine in clinical practice.
    Molecular pathogenesis is not fully elucidated.

Cost of drug development
    Comparison with standard of care requires larger studies
      to demonstrate clinical benefit.
    Proof of cardiovascular safety demands costly and complex
      trials.
    Impact on diabetes complications takes too long to achieve.
    The multiplicity of available pharmacological options
      constrains the therapeutic niche for novel agents,
      undermining viability.

Inadequacy of current practices
    Preclinical models may not be relevant to the human
      situation.
    Modulating glycemia may not be the critical end point.
    Emergence of side effects in humans threatens new
      agents, as hyperglycemia does not confer immediate
      serious risk and can be controlled via other means.
      Initial evaluation of these side effects in phase 1 and
      2 trials may be inefficient, insufficient, and expensive.

What can one do to understand pathophysiology better, aiming to identify the key molecular targets that will subserve the production of disease-modifying drugs, so that these can be prescribed to the patient who harbors the corresponding disease subtype? Can we improve our methods for target validation in the relevant model system, the human, ahead of costly and risky clinical testing? Can we enhance our predictive abilities around efficacy and safety before we launch the necessary definitive clinical trials?

In this Perspective, I will use the vantage point of type 2 diabetes to argue that unbiased genetic discovery in humans can indeed support these efforts, identify valid drug targets, illuminate mechanisms, flag off-target effects, and provide causality. The hope is that facilitating the deployment of new genetic knowledge across pharmaceutical discovery programs will accelerate drug development by enhancing the efficiency and cost-effectiveness of bringing new agents to market.

MODERN GENOMIC DISCOVERY

The sequencing of the human genome, the characterization of the patterns of human genetic variation, and technological and methodological advances in genotyping and sequencing studies have underwritten a veritable explosion in genetic discovery (10). Crucially, these studies have queried the entire human genome in an agnostic fashion, free from the constraints of preexisting biological knowledge, thus enabling the implication of heretofore unsuspected pathways. Larger sample sizes achieved via international collaboration, improved imputation methods, and next-generation sequencing techniques have expanded the allele frequency spectrum for variant association, allowing for the detection of low-frequency variants and the targeting of specific ethnic subgroups (Table 2). In this manner, over the past decade, nearly 100 loci have been associated with type 2 diabetes or related traits in multiple populations (Fig. 1) (11). Though together these variants only explain 10–15% of the inherited cause of type 2 diabetes, the approach has proven successful and the methods have been streamlined. It is likely that the accrual of larger sample sizes (e.g., in developing nations or large health care systems) as costs continue to drop will only continue to advance discovery.

Have these genomic studies generated new knowledge? For the purposes of drug target identification in type 2 diabetes, several key insights have emerged. Genome-wide association studies (GWAS) have established β-cell function as the focus in type 2 diabetes pathogenesis, complementing prior observations in monogenic diabetes (12). They have revealed causal links between metabolism and circadian rhythmicity, fetal development, or lipid regulation that were previously highlighted by epidemiological correlations (13). They have identified new pathways (e.g., zinc transport into β-cell granules [14], KLF14 target genes in adipocytes [15], melatonin signaling [16], or monocarboxylate transport [17]) in type 2 diabetes pathogenesis. They have also enabled a more comprehensive

Table 2—Types of genetic studies

Type	Alleles captured	Advantages	Limitations
Targeted genotyping	Specific variants	Inexpensive, hypothesis driven	Constrained by current knowledge, cannot use genome to control for population effects
Genome-wide genotyping (GWAS)	Common; coding and noncoding	Affordable, comprehensive, agnostic, can control for population effects, streamlined analysis	Requires large sample sizes to detect modest effects at genome-wide statistical significance ($P = 5 \times 10^{-8}$)
Exome-wide genotyping	Common and low-frequency; coding	Affordable, comprehensive as far as genes are concerned, agnostic, can control for population effects, can conduct individual variant testing as well gene burden tests, easier interpretation of functional effects	Requires large sample sizes to detect modest effects at exome-wide statistical significance ($P = 5 \times 10^{-7}$ for single variants, $P = 2.5 \times 10^{-6}$ for gene-based tests of rare variant aggregation), only focuses on coding variation that is shared across populations
Whole-exome sequencing	Common, low-frequency, and rare; coding	Expensive; comprehensive as far as genes are concerned; agnostic; can control for population effects; can conduct individual variant testing as well gene burden tests; can discover novel variants in an individual, a family, or a group; easier interpretation of functional effects	Requires large sample sizes to detect modest effects at exome-wide statistical significance ($P = 5 \times 10^{-7}$ for single variants, $P = 2.5 \times 10^{-6}$ for gene-based tests of rare variant aggregation), capture of variation may be uneven across the genome
Whole-genome sequencing	Common, low-frequency, and rare; coding and noncoding	Very expensive, most comprehensive, agnostic, can control for population effects, can discover novel variants in an individual, a family, or a group	Unresolved threshold for statistical significance in the low-/rare frequency spectrum, challenging interpretation of functional effects

exploration of the genetic architecture of the disease, setting boundaries for the effect sizes and allelic series that make up the likely universe of disease-causing variation (18).

The picture that emerges from the empirical evidence is one by which several hundred to a few thousand genetic variants of very modest effects are likely to seed the genetic predisposition to type 2 diabetes, interacting with a multitude of environmental insults. Given the number of contributing factors involved and the weak effect of any individual determinant, the definition of subtypes is unlikely to be as cleanly demarcated as it is for monogenic disease; instead, it may have to rely on drawing somewhat arbitrary lines along various continua that are genetically and/or physiologically defined, denoting distinct extremes along axes of pathophysiology. To borrow Mark McCarthy's analogy, the challenge will be to describe specific hues across the spectra of a multicolored palette (19).

GENETICALLY DRIVEN DIABETES SUBCLASSIFICATION

Under this paradigm, have genetic findings improved type 2 diabetes nosology? As the number of genetic associations reaches critical mass and new associations emerge from parallel genomic studies for related phenotypes, investigators can use a number a clustering approaches to group genomic loci around select limbs of the glucose homeostasis system. In an early exploration, type 2 diabetes–associated loci could be subdivided into clusters that impair β-cell function or insulin sensitivity (20). A more focused effort, centered on variants associated with insulin resistance, demonstrated that a subset of such variants defined a lipodystrophy-like

syndrome (21): a genetic risk score (GRS) constructed with 11 insulin resistance–raising variants was associated with lower BMI but higher risk of type 2 diabetes, nonalcoholic fatty liver disease, hypertension, and coronary artery disease. The growing list of genetic associations, larger sample sizes, and richer phenotypic data sets will only continue to clarify the existence of subgroups that can be defined by extremes in a range of such GRSs, such that the clinical approach to their surveillance and treatment can be tailored more rationally.

The use of GRSs is needed to improve statistical power in capturing a larger proportion of the variance in any given trait because of the modest effects exerted by individual genetic variants. However, there are instances where a single association is sufficient for decision making. Typically this happens in the context of rare or low-frequency variants that have strong effects in specific populations. A nonsense polymorphism in *TBC1D4* has a 17% minor allele frequency in Inuit populations, raises 2-h glucose, and increases type 2 diabetes risk 10-fold (22). As TBC1D4 is implicated in transducing the insulin signal in skeletal muscle, it is believed that these individuals suffer from a type 2 diabetes mostly defined by muscle insulin resistance and might benefit preferentially from treatment with an insulin sensitizer (23). Similarly, a missense polymorphism in *HNF1A* has a 2% minor allele frequency in Latino populations and increases type 2 diabetes risk fivefold (24). Because carriers of loss-of-function mutations in this gene experience a more favorable response to sulfonylureas, it is possible that these patients might be better treated with those agents as well, at least early in their disease course.
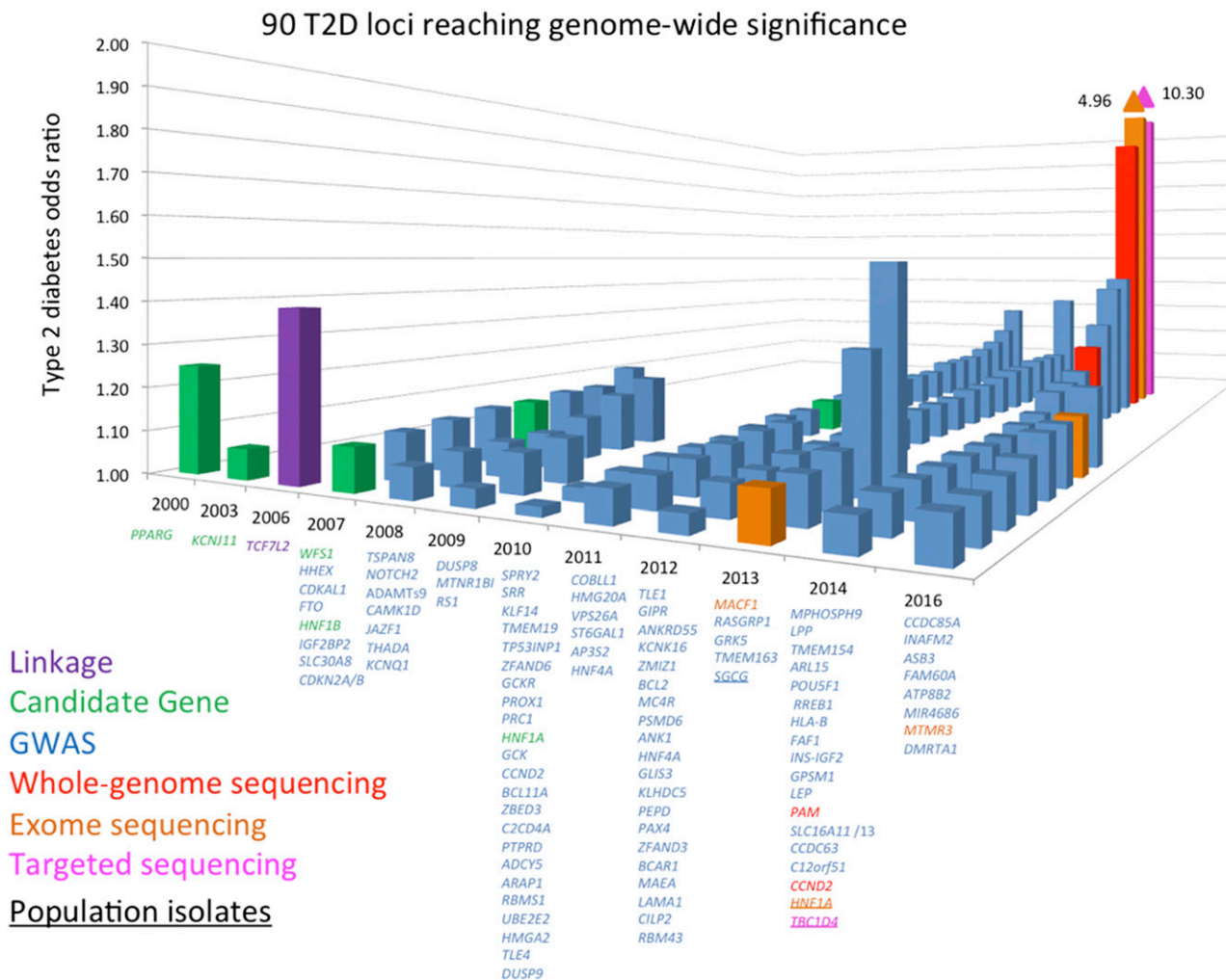
Figure 1—Chronological listing of type 2 diabetes–associated loci, plotted by year of definitive publication and approximate effect size. They are named by the nearest gene, though this convention does not indicate that the causal gene has been found at the locus. Candidate loci are shown in green, loci discovered via agnostic genome-wide association approaches in blue, loci identified by exome sequencing in orange, and loci identified by whole-genome sequencing in red. *TCF7L2* (shown in purple) was discovered by dense fine-mapping under a linkage signal. *TBC1D4* (shown in pink) was identified by exome sequencing of a locus found to be associated with a diabetes-related quantitative trait. Gene names that are underlined denote identification in population isolates. T2D, type 2 diabetes.

AGNOSTIC GENOMIC STUDIES CAN YIELD DRUG TARGETS

Is this knowledge relevant to drug discovery? There are several ways of answering this very pertinent question (Table 3). One can ask whether genetic studies have uncovered true positive findings, i.e., instances where a known drug target is encoded by a gene detected via these methods. This would add confidence that the approach is effective. As a higher burden of proof, one can demand to see examples where genetic studies have led to the development of successful drugs approved for use in patients. Through the different lens of the existing pharmacopeia, one can ask whether the genes that encode approved drug targets are enriched for type 2 diabetes–associated variants. And finally, one can also ask whether genetic studies can shed light on the drug targets of currently approved agents when these remain obscure.

Indeed, genetic association studies for type 2 diabetes and fasting glucose have detected variants in genes that encode existing drug targets: *PPARG* for thiazolidinediones (25), *KCNJ11* for sulfonylureas (26), and *GLP1R* for GLP-1 receptor agonists (27). In related fields, a noncoding variant in the *HMGCR* gene (encoding HMG-CoA reductase) has a small effect on LDL cholesterol, but it flags this gene as a valid target for therapeutic development (28). In other words, if nothing had been known about thiazolidinediones, sulfonylureas, GLP-1 receptor agonists, or cholesterol biosynthesis prior to the onset of GWAS, these studies would have pointed to these genes as potential targets for therapeutic design. These findings also illustrate that the modest effects generated by a comparison of allele frequencies of common variants in these loci between case and control subjects do not undermine the likelihood that the genes, molecules, or pathways

Table 3—Evidence of utility of genetic approaches in drug target identification in type 2 diabetes and related traits

Retrospective: Genetic studies have yielded associated genes that are known targets for currently marketed medications.

Prospective: Genetic studies (in Mendelian disease) have yielded target genes for which novel drugs have been developed and approved.

Genes that encode existing drug targets are enriched for variants that are associated with type 2 diabetes.

Unbiased genomic searches can uncover loci associated with drug response.

revealed by these approaches can serve as viable therapeutic targets.

Similarly, genetic studies in other related diseases have paved the way for the introduction of successful therapies. Knowledge about impaired cellular trafficking of the cystic fibrosis transmembrane regulator led to the development of ivacaftor and lumacaftor, transformative therapies for cystic fibrosis (29,30). Identification of healthy carriers of loss-of-function *PCSK9* mutations ushered PCSK9 inhibition as a novel approach in LDL lowering (31,32), and characterization of families who had lost SGLT2 function enabled the introduction of SGLT2 inhibitors as the most recent type 2 diabetes drug class (33). In polygenic disease, this proof has been more laborious to attain, partly because of the relatively early state of the field.

Nevertheless, our group has mined GWAS to determine whether genes that encode the targets for approved type 2 diabetes drugs are enriched for type 2 diabetes–associated variants (34). We compiled a list of 102 genes in pathways targeted by available antihyperglycemia medications and applied a new statistical method modified from transcriptomic analyses to ascertain whether this gene set was enriched for type 2 diabetes genetic associations. This was indeed the case (at a highly significant P value of 2×10^{-5}) and was independently replicated. The approach can also be used to unmask potential side effects by mining GWAS for other traits.

Finally, pharmacogenetic studies can be used to search for the unknown targets of existing agents. In type 2 diabetes, the most tantalizing example concerns metformin, the first-line therapy in all treatment algorithms (3). Finding its molecular target has proven elusive. Although a number of pathways have been shown to be modulated by metformin action (including mitochondrial complex I [35], AMPK [36], cyclic AMP [37], mitochondrial glycerophosphate dehydrogenase [38], and, more recently, the nuclear pore complex [39]), its precise molecular target is not known. By leveraging cohorts where DNA is available and metformin response can be quantified, GWAS can begin to identify genomic loci that are associated with metformin response (40,41) and harbor genes responsible for the observed effects.

FROM GENETIC ASSOCIATION TO EFFECTOR TRANSCRIPT

Confirming robust genomic associations is only the beginning. These signals serve to plant a flag in a given genomic region, where a haplotype (a linear arrangement of correlated genetic variants) is more often present in disease than in health. However, the physical proximity of the index variant to a protein-coding gene does not imply that this is the gene that, when mutated, gives rise to the phenotype. The variant could be disrupting an enhancer element or another regulatory region for more distant genes (including those that encode microRNAs or long noncoding RNAs, for instance), misleading naive investigators about the relevant drug target. Thus, it is essential that genomic studies be followed by principled searches for the effector transcript that underlies each genetic association.

One potential avenue involves the discovery of coding mutations that disrupt protein function and phenocopy the original association. Typically these are less well tolerated and therefore present at lower allele frequencies. Exome genotyping or sequencing studies are required to detect them in high enough numbers to derive convincing statistical confidence (Table 2). Coding variants can also be aggregated into gene burden tests to increase statistical power (42). When present, they provide supportive evidence that the original GWAS association marked the gene where they lie as the likely effector transcript. Ancillary information on the pattern of tissue expression of index genes can be found in the Genotype-Tissue Expression (GTEx) database, which combines expression and human genomic data across many human tissues (43,44). This allows one to establish the presence of the transcript of interest in physiologically relevant organs and examine whether noncoding variants associated with the disease phenotype affect message levels (expression quantitative trait loci [eQTL] analysis). Experimental validation that the allelic change leads to the expected perturbation in enhancer or promoter activity is arduous to obtain but no less crucial in demonstrating causality.

Identifying a likely effector transcript via the above approaches does not by itself establish the direction of effect. That is, even a missense mutation that is associated with a disease phenotype at genome- or exome-wide statistical significance does not per se indicate whether the disease-associated allele induces gain or loss of function at the molecular level. Indeed, the amino acid change may impair or enhance the activity of an enzyme, transporter, or transcription factor, and either one of the two actions could lead to metabolic dysregulation at the organismal level. Additional information is required for the pharmaceutical industry to launch an experimental program based on that putative drug target, as the search, design, and evaluation of activators or inhibitors might be radically different depending on which avenue is selected.

Genetic analyses can guide this decision. At times, variants that change amino acid sequence will have a clear

effect on protein function, aligning the direction of the molecular consequence with the disease risk allele. Very often, however, a single amino acid change has no discernible impact, and a search for mutations that alter the protein unambiguously becomes necessary. Through large-scale sequencing approaches, investigators with access to diverse cohorts can identify protein-truncating variants (PTVs) that disrupt protein function (e.g., stop codons, intron-exon splice acceptor sites, frameshifts, or read-through mutations), enabling the study of physiological consequences of haploinsufficiency at that site in living humans. If PTVs are statistically more frequent in disease than in health, it can be presumed that their effect on the protein (whether loss of function by deletion of a key activity domain or gain of function by deletion of an inhibitory domain) is deleterious, and therapies should counteract this effect by either raising the activity or expression of the affected protein (if the PTV induces loss of function) or inhibiting its activity or expression (if the PTV induces gain of function). The reciprocal strategies would be applied if PTVs are found to be protective. An elegant illustration of this concept was rendered by the observation that loss of function at the zinc transporter encoded by *SLC30A8* appears to be protective for type 2 diabetes, clarifying the direction of effect of the index R235W coding variant (45). Finally, corroborating proof can be obtained by overexpression, silencing, or knockout experiments in appropriate cellular or animal model systems, now facilitated by genome editing technologies such as CRISPR-Cas9. It should be kept in mind that although supportive evidence is helpful, the absence of a consistent effect in experimental models does not by itself undermine the human genetic associations, as the effects could be species specific or require the interaction of multiple organ systems.

FROM EFFECTOR TRANSCRIPT TO DRUG

Once geneticists, human physiologists, and experimentalists have zeroed in on a valid target, the drug development team must establish the general druggability of the target, that is, how likely it is that a small molecule or biologic designed to target the gene will do so successfully and generate the desired therapeutic effect. Several considerations influence that assessment. First, at what developmental stage does the biological effect that must be perturbed occur? If the damage takes place early in development, e.g., by establishing a ceiling for a person's β-cell mass in utero, intervening therapeutically may be challenging. Second, where in the body is the gene expressed, i.e., what other organs might be affected by systemic delivery of the drug? Third, where is the protein or RNA product localized (secreted into the circulation or on the cell surface, embedded in the plasma membrane, untethered in the cytosol, inside a specific organelle, or in the nucleus)? Fourth, what are the three-dimensional constraints that determine whether a small molecule will be able to interfere with the protein function? Last, how

specific will that perturbation be, in terms of possible off-target effects on related molecules or in other tissues? When all of these issues are weighed, only a handful of proven drug targets may emerge as sufficiently attractive to invest the sizable human and technological resources and temporal and financial commitments required for a serious drug program to enjoy a decent chance of success.

Once again, human genetics may facilitate some of these necessary evaluations. With respect to off-target effects that could render a therapeutic candidate unsafe for use in humans, investigators can use available databases to gauge the likelihood that disrupting a given gene may cause untoward side effects. If loss-of-function carriers exist and these are free of a discernible clinical phenotype, one can be reasonably assured that interfering with that gene product's function may be safe. This does not preclude the conduct of appropriate preclinical, phase 1, or phase 2 studies, as it is possible that a permanent loss of function from the time of conception may induce compensatory pathways to overcome the genetic defect, although these may not be plastic enough for their unfolding at a more advanced developmental stage to defend against a loss of function imposed later in life. Nevertheless, a benign clinical phenotype of mutation carriers may provide assurances that investing in this program is worthwhile. The assembly of large numbers of protein-coding variation by the Exome Aggregation Consortium (ExAC) (46,47) and its successor Genome Aggregation Database (gnomAD) is one way to streamline this task. Interrogating electronic medical records paired with genomic information by health care systems, such as Kaiser Permanente, Geisinger, the UK Biobank, Mount Sinai's BioMe, Vanderbilt's BioVU, and others in the Electronic Medical Records and Genomics (eMERGE) Network, allows investigators to determine whether carriage of specific variants is associated with unrelated clinical diagnoses.

GENETICS AND CAUSAL INFERENCE

The use of genomic data to identify drug targets leverages a unique advantage of the genetic approach: alone among all biomarkers, inherited genetic variation always precedes the disease process and is unaffected by it or by its treatment. Thus, in contrast to epigenomics, transcriptomics, metabolomics, or proteomics, it is not susceptible to reverse causation. It is still vulnerable to limited confounding, for example, if the disease prevalence varies by ethnic groups and the associated allele is a marker of ethnicity rather than disease, but this type of confounding (caused by population stratification) can be easily controlled by harnessing the rest of the genome, presumably unrelated to disease, in providing the necessary statistical adjustments.

This exceptional feature of genetic approaches can also be used to support drug discovery programs. Epidemiological observations may have suggested that a particular biomarker is correlated with pathology, and longitudinal studies may have indicated that levels of said biomarker

rise in anticipation of disease onset. A reasonable assumption can be made that modulating the biomarker may have an impact on disease incidence. However, it is still entirely possible that the biomarker may be an epiphenomenon, driven by an occult primary process that causes disease and elevates the biomarker in parallel, whereas the biomarker itself has no direct influence on pathogenesis.

Until recently, to address the potential causal role of the biomarker, pharmaceutical companies had to produce the means to modulate biomarker levels and test whether such modulation affected disease outcomes in randomized clinical trials. Now genetics can aid in this high-stakes decision making (Fig. 2). Because alleles are randomized at meiosis, lifelong exposure to a genetic variant is largely a random event. If a variant raises levels of a biomarker and that biomarker is causal for disease, then—contingent on adequate statistical power—the biomarker-raising allele should be associated with the disease outcome. If, however, despite clear effects on biomarker levels, there is no hint of an association with disease, then merely modulating levels of the biomarker may have no influence on the disease process. This technique, termed Mendelian randomization (48), has been used to demonstrate that LDL cholesterol is causal for myocardial infarction (as had been demonstrated by multiple statin trials), whereas HDL cholesterol is not (as corroborated by failed HDL-raising randomized clinical trials, conducted at tremendous expense and effort) (49). In a similar fashion and through the use of GRSs, we have recently demonstrated that BMI influences diabetic kidney disease in type 1 diabetes (50) and hyperglycemia is causal for coronary artery disease (51).

DEMOCRATIZING GENETICS

The sheer size and complexity of genetic analyses have often conspired to make genetic data sets only accessible to the cognoscenti. Without a background in statistical
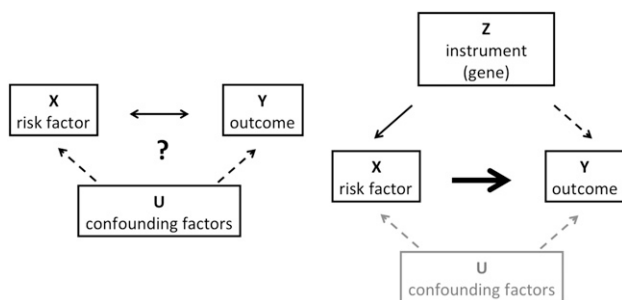
Figure 2—Schema illustrating the concept of Mendelian randomization. Left: A risk factor X is observed to co-occur with a clinical outcome Y. The relationship between the two is unclear, as the risk factor could cause the outcome, be caused by it, or both could be driven by occult confounding factors U. Right: However, if a genetic instrument Z is found that determines levels of the risk factor and meets a number of assumptions (e.g., no pleiotropy), then detecting an association of the instrument with the outcome (dashed arrow) places the risk factor on the causal pathway for the outcome (bold arrow).

genetics or bioinformatics, it has been very difficult for interested parties in academia, government, or industry to engage genetic data sets to answer critical questions. Thus, emergent findings in human genetics have not truly permeated the rest of biology, and experimentalists have been largely unable to test biological hypotheses anchored on human genetic data.

Human geneticists have become aware of this challenge. Though typically attuned to the ethical imperative of data sharing, as manifested by the commonly accepted standard in the field of making summary data publicly available via consortium websites, they have found the official mechanisms available for such sharing imposing, burdensome, and inadequate. To overcome these barriers, the Accelerating Medicines Partnership in Type 2 Diabetes (AMP-T2D), involving government, industry, and academia, has coalesced to create a knowledge portal (www.type2diabetesgenetics.org) where genetic and phenotypic information around type 2 diabetes and related traits will be deposited for data mining (52). The database, populated by existing genomics consortia for type 2 diabetes (DIAbetes Genetics Replication And Meta-analysis [DIAGRAM] and Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples [T2D-GENES]), quantitative glycemic traits (Meta-Analyses of Glucose and Insulin-related traits Consortium [MAGIC]), *trans*-ethnic explorations (Meta-Analysis of Type 2 Diabetes Genome-Wide Association Studies in African Americans [MEDIA], African American Genetics of Glucose and Insulin [AAGILE], Slim Initiative in Genomic Medicine for the Americas [SIGMA], and DIAbetes Meta-ANalysis of Trans-Ethnic association studies [DIAMANTE]), and diabetes complications (GEnetics of Nephropathy: an International Effort [GENIE] and Diabetic Nephropathy Collaborative Research Initiative [DNCRI]), health care organizations (e.g., Mount Sinai's BioMe), pharmaceutical companies (e.g., CArdiovascular and Metabolic Patient cohort [CAMP], sponsored by Pfizer), and many others, attempts to capture the majority of genomic information available globally for this condition. It resides in a secure set of sites linked to each other via federation. Analytical engines are being developed that allow the user to query the data sets with intelligent and flexible requests in real time. To protect research participants, only summary results will be returned, and no primary data can be downloaded. The analytical interface is modular, versatile, and organic, adopting new methods and perspectives while preserving rigor.

CONCLUSIONS

There is a need for a revolution in drug discovery in type 2 diabetes, with the twin goals of disease modification and alleviation of specific pathophysiological processes. Biologists, epidemiologists, and physiologists must collaborate in defining clear disease subtypes. The focus must switch to the human as the relevant model system. Genetics can help clarify disease heterogeneity and provide valid candidates for drug development. Placing genotype and phenotype

Table 4—Human genetic applications in drug discovery

Detection of genomic regions associated with the phenotype of interest

Evaluation of strength of association of the same region with endophenotypes, related traits, or other clinical outcomes

Fine-mapping of the region to focus on the likely causal variant

Assessment of coding variation or eQTL in relevant tissues to identify the causal transcript

Study of protein-truncating variants to determine direction of effect

Integration of other genomic data to explore potential off-target effects

Use of Mendelian randomization to establish causality

information in a secure, accessible, user-friendly, and comprehensive site such as the AMP-T2D Knowledge Portal is one initial step in that direction. Robust and intelligent genetic analyses can provide shortcuts that identify effector transcripts in genomic regions, establish direction of functional effect, support causal inference around intermediate biomarkers, and illustrate off-target consequences (Table 4). The rational and comprehensive deployment of genetic approaches across the pharmaceutical industry should accelerate and enhance drug discovery pipelines in type 2 diabetes.

References

1. Cefalu WT. Pharmacotherapy for the treatment of patients with type 2 diabetes mellitus: rationale and specific agents [published correction appears in Clin Pharmacol Ther 2007;81:910]. Clin Pharmacol Ther 2007;81:636–649

2. Kahn SE, Cooper ME, Del Prato S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. Lancet 2014;383:1068–1083

3. Inzucchi SE, Bergenstal RM, Buse JB, et al. Management of hyperglycemia in type 2 diabetes, 2015: a patient-centered approach: update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes. Diabetes Care 2015;38:140–149

4. Skyler JS, Bakris GL, Bonifacio E, et al. Differentiation of diabetes by pathophysiology, natural history, and prognosis. Diabetes 2017;66:241–255

5. Rosenblatt M. The large pharmaceutical company perspective. N Engl J Med 2017;376:52–60

6. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov 2012;11: 191–200

7. Dieleman JL, Baral R, Birger M, et al. US spending on personal health care and public health, 1996-2013. JAMA 2016;316:2627–2646

8. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol 2014;32:40–51

9. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. Nat Rev Drug Discov 2013;12:581–594

10. Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med 2010;363:166–176

11. Mohlke KL, Boehnke M. Recent advances in understanding the genetic architecture of type 2 diabetes. Hum Mol Genet 2015;24(R1):R85–R92

12. Florez JC. Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? Diabetologia 2008;51:1100–1110

13. Billings LK, Florez JC. The genetics of type 2 diabetes: what have we learned from GWAS? Ann N Y Acad Sci 2010;1212:59–77

14. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881–885

15. Small KS, Hedman AK, Grundberg E, et al.; GIANT Consortium; MAGIC Investigators; DIAGRAM Consortium; MuTHER Consortium. Identification of an imprinted master trans regulator at the *KLF14* locus related to multiple metabolic phenotypes. Nat Genet 2011;43:561–564

16. Tuomi T, Nagorny CL, Singh P, et al. Increased melatonin signaling is a risk factor for type 2 diabetes. Cell Metab 2016;23:1067–1077

17. Williams AL, Jacobs SB, Moreno-Macías H, et al.; SIGMA Type 2 Diabetes Consortium. Sequence variants in *SLC16A11* are a common risk factor for type 2 diabetes in Mexico. Nature 2014;506:97–101

18. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. Nature 2016;536:41–47

19. McCarthy MI. Painting a new picture of personalised medicine for diabetes. Diabetologia 2017;60:793–799

20. Dimas AS, Lagou V, Barker A, et al.; MAGIC Investigators. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes 2014;63:2158–2171

21. Yaghootkar H, Scott RA, White CC, et al. Genetic evidence for a normal-weight "metabolically obese" phenotype linking insulin resistance, hypertension, coronary artery disease, and type 2 diabetes. Diabetes 2014;63:4369–4377

22. Moltke I, Grarup N, Jørgensen ME, et al. A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes. Nature 2014;512: 190–193

23. Manousaki D, Kent JW Jr, Haack K, et al. Toward precision medicine: TBC1D4 disruption is common among the Inuit and leads to underdiagnosis of type 2 diabetes. Diabetes Care 2016;39:1889–1895

24. Estrada K, Aukrust I, Bjørkhaug L, et al.; SIGMA Type 2 Diabetes Consortium. Association of a low-frequency variant in *HNF1A* with type 2 diabetes in a Latino population [published correction appears in JAMA 2014;312:1932]. JAMA 2014;311:2305–2314

25. Altshuler D, Hirschhorn JN, Klannemark M, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet 2000;26:76–80

26. Gloyn AL, Weedon MN, Owen KR, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes 2003;52:568–572

27. Wessel J, Chu AY, Willems SM, et al.; EPIC-InterAct Consortium. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun 2015;6:5897

28. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet 2009;41:56–65

29. Wainwright CE, Elborn JS, Ramsey BW, et al.; TRAFFIC Study Group; TRANSPORT Study Group. Lumacaftor–ivacaftor in patients with cystic fibrosis homozygous for Phe508del CFTR. N Engl J Med 2015;373:220–231

30. Ramsey BW, Davies J, McElvaney NG, et al.; VX08-770-102 Study Group. A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. N Engl J Med 2011;365:1663–1672

31. Robinson JG, Farnier M, Krempf M, et al.; ODYSSEY LONG TERM Investigators. Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. N Engl J Med 2015;372:1489–1499

32. Sabatine MS, Giugliano RP, Wiviott SD, et al.; Open-Label Study of Long-Term Evaluation against LDL Cholesterol (OSLER) Investigators. Efficacy and safety of evolocumab in reducing lipids and cardiovascular events. N Engl J Med 2015;372:1500–1509

33. Ferrannini E, Solini A. SGLT2 inhibition in diabetes mellitus: rationale and clinical prospects. Nat Rev Endocrinol 2012;8:495–502

34. Segrè AV, Wei N, Altshuler D, Florez JC; DIAGRAM Consortium; MAGIC Investigators. Pathways targeted by antidiabetes drugs are enriched for multiple genes associated with type 2 diabetes risk. Diabetes 2015;64:1470–1483

35. Owen MR, Doran E, Halestrap AP. Evidence that metformin exerts its anti-diabetic effects through inhibition of complex 1 of the mitochondrial respiratory chain. Biochem J 2000;348:607–614

36. Zhou G, Myers R, Li Y, et al. Role of AMP-activated protein kinase in mechanism of metformin action. J Clin Invest 2001;108:1167–1174

37. Miller RA, Chu Q, Xie J, Foretz M, Viollet B, Birnbaum MJ. Biguanides suppress hepatic glucagon signalling by decreasing production of cyclic AMP. Nature 2013;494:256–260

38. Madiraju AK, Erion DM, Rahimi Y, et al. Metformin suppresses gluconeo-genesis by inhibiting mitochondrial glycerophosphate dehydrogenase. Nature 2014;510:542–546

39. Wu L, Zhou B, Oshiro-Rapley N, et al. An ancient, unified mechanism for metformin growth inhibition in *C. elegans* and cancer. Cell 2016;167:1705–1718.e13

40. Zhou K, Bellenguez C, Spencer CC, et al.; GoDARTS and UKPDS Diabetes Pharmacogenetics Study Group; Wellcome Trust Case Control Consortium 2; MAGIC Investigators. Common variants near *ATM* are associated with glycemic response to metformin in type 2 diabetes. Nat Genet 2011;43:117–120

41. Zhou K, Yee SW, Seiser EL, et al.; MetGen Investigators; DPP Investigators; ACCORD Investigators. Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin. Nat Genet 2016;48:1055–1059

42. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 2011;89:82–93

43. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580–585

44. Melé M, Ferreira PG, Reverter F, et al.; GTEx Consortium. Human genomics. The human transcriptome across tissues and individuals. Science 2015;348:660–665

45. Flannick J, Thorleifsson G, Beer NL, et al.; Go-T2D Consortium; T2D-GENES Consortium. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. Nat Genet 2014;46:357–363

46. Lek M, Karczewski KJ, Minikel EV, et al.; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285–291

47. Karczewski KJ, Weisburd B, Thomas B, et al.; The Exome Aggregation Consortium. The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Res 2017;45:D840–D845

48. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 2008;27:1133–1163

49. Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet 2012;380:572–580

50. Todd JN, Dahlström EH, Salem RM, et al.; FinnDiane Study Group. Genetic evidence for a causal role of obesity in diabetic kidney disease. Diabetes 2015;64:4238–4246

51. Merino J, Leong A, Posner DC, et al. Genetically driven hyperglycemia increases risk of coronary artery disease separately from type 2 diabetes. Diabetes Care 2017;40:687–693

52. Flannick J, Florez JC. Type 2 diabetes: genetic data sharing to advance complex disease research. Nat Rev Genet 2016;17:535–549